**3.1 随机变量和离散分布** 2020年1月19日09点34分

**例题3.1.1** 抛硬币 抛10次公平硬币,将正面出现的次数定义为随机变量,是样本空间中元素的变量.

**定义3.1.1随机变量** 设是某项实验的样本空间.定义在上的实数函数被称为随机变量.

**例题3.1.2** 测量人体身高 这里身高就是随机变量

**例题3.1.3** 资源需求 举例随机变量,这里对随机变量的定义稍微复杂

**随机变量的分布**

当某项实验的样本空间的概率已经测得,那么我们就可以确定每一个随机变量可能值的概率.设是实数子集使得是一个事件,设表示的值属于子集的概率.则等于实验结果使得的概率.用符号表示为

**定义3.1.2 分布** 随机变量的分布是形如所有概率的集合,其中是任意实数子集并使得是一个事件.

**例题3.1.4** 抛硬币一般化

**例题3.1.3** 资源需求一般化

**定义3.1.3 离散分布/随机变量** 如果随机变量只能取个不同值或至多无限个不同值.则我们称有离散分布或是离散随机变量.

**定义3.1.4 概率函数/p.f./支持** 如果随机变量有离散分布,则的概率函数(简称p.f.)定义为函数使得对于每一个实数,

闭集被称为的支持(分布).

注意:概率函数有时在其它教材被称为概率质量函数,或p.m.f.但是在本书中我们不使用此术语.

**例题3.1.6** 资源需求 对**例题3.1.3** 提出的随机变量的解答.

**例题3.1.7** 抛硬币

**定理3.1.1** 设随机变量的p.f.为.如果不是的可能值,则.同样,如果序列包含所有可能值,则.

**定理3.1.2** 设有离散分布,实线的每一个子集的概率可以从下列关系确定

**定义3.1.5 伯努利分布/随机变量** 随机变量只能取两个值0和1,且,则被称为参数为的伯努利随机变量.

**例题3.1.7** 抽奖号码

**定义3.1.6** 整数的均匀分布 设是整数.假设随机变量的值等可能地出现在区间上.则我们称为在区间上有均匀分布.

**定理3.1.3** 如果在区间上有均匀分布,则的p.f.是

**例题3.1.9** 零件损坏 假设一台机器生产的零件损坏率为,并假设这台机器生产了个零件,设随机变量是零件损坏的个数,显然的取值范围是区间.则的p.f.为

**定义3.1.7 二项分布/随机变量** 由公式所表达的离散分布被称为*参数是和的二项分布*.符合该分布的随机变量被称为*参数是和的二项随机变量*.

**3.2 连续分布** 2020年1月19日18点14分

**例题3.2.1** 资源需求

**定义3.2.1 连续分布/随机变量** 如果存在一个定义在实数上的函数,对于实数任意区间(开区间或闭区间均可),使得随机变量在该区间上取值的概率等于在该区间的积分,则我们称有连续分布或是连续随机变量.

**定义3.2.2 概率密度函数/p.d.f./支持** 如果有连续分布,则定义3.2.1所描述的函数被称为的概率密度函数(简称p.d.f.).闭集被称为的支持(分布).

根据上述定义,对于任意闭区间

每一个p.d.f.必须满足下列两个需求:

和

**注意:连续分布为单个值赋予的概率为0**.公式与和是相等的.因此,从连续分布的定义得到,如果有连续分布,对于每一个值,.但是并不意味着不会发生.

**例题3.2.2** 温度预报 引出下列定义

**定义3.2.3 区间上的均匀分布** 设和是给定的两个实数并且.设是随机变量且已知,对于的每一个子区间,属于给定区间的概率与该区间的长度成正比.我们则称随机变量*在区间上有均匀分布*.

**定理3.2.1** 设在区间上有均匀分布,则的p.d.f.为

**注意**:**密度不等于概率**.p.d.f.是无界限的;在处的p.d.f.无法给出在附近的概率,只能通过p.d.f.在附近的积分得到;由于无界区间的长度是无穷的,因此无法在无界区间定义均匀分布;p.d.f.可以定义在开区间和半开区间,并且定义相同,例如公式定义在上与定义在上是相同的.

**例题3.2.3** 不完整的p.d.f.

**注意:计算归一化常(系)数**.某些随机变量是无界的，如果仅仅计算有界范围内的概率分布，则p.d.f.需要除有界范围内的总概率.

**例题3.2.4** 计算p.d.f.

**例题3.2.5** 无界随机变量

**例题3.2.6** 无界p.d.f.

**例题3.2.7** 混合p.d.f. 对**例题3.2.5**的扩展

**3.3 累积分布函数** 2020年2月5日14点17分

**例题3.3.1** 电压 是**例题3.2.5**的延续

**定义3.3.1** **累积分布函数** 随机变量的分布函数或累积分布函数(缩写为c.d.f.)定义为

**例题3.3.2** 伯努利c.d.f.设为参数为的伯努利随机变量,即.则的c.d.f.为

**属性3.3.1** **非递减** 随着增大是非递减的;即,如果,则.

**属性3.3.2** 且.

**属性3.3.3** 右侧连续 如果满足任意值,则c.d.f.总是在右侧连续.

**例题3.3.3** 电压 是**例题3.3.1**的延续

**定理3.3.1** 对每一个值,

**定理3.3.2** 对所有的值和值使得,

**定理3.3.3** 对每一个值,

**定理3.3.4** 对每一个值,

**定理3.3.5** 设是连续分布,并且设和分别表示其p.d.f.和c.d.f.则在每一个处都连续,

并且

在所有上使得连续.

**例题3.3.4** 从c.d.f.计算p.d.f.

**例题3.3.5** 预测降雨量,引出分位数函数

**定义3.3.2 分位数/百分位数** 设是一个随机变量,其c.d.f.为.对于每一个严格位于0和1之间的值,定义是使得的最小值.则被称为**分位数[quantile]**或的100**百分位数[percentile]**.函数定义在开区间上被称为的**分位数函数[quantile function]**.

**例题3.3.6** 分位数函数的应用

**连续分布的分位数** 当随机变量的c.d.f.是连续的并且在的所有可能值的集合中一对一,则的反函数存在并且等于的分位数函数.

**例题3.3.7** 投资风险 分位数函数的应用,**没太认真看**

**例题3.3.8** 均匀分布的分位函数

**例题3.3.9** 二项分布的分位数

**定义3.3.3 中位数/分位数** 分布的1/2分位数或第50个百分位数称为其中位数.1/4分位数或25个百分点是下四分位数.3/4分位数或第75个百分位数称为上四分位数.

**3.4 二元分布** 2020年2月9日09点58分

**例题3.4.1 例题3.1.5**引申,水电需求问题

**定义3.4.1 联合/二元分布** 设和是随机变量.和的联合分布或二维分布是形如的所有实数对使得是一个事件的集合的概率.

**例题3.4.2** 剧院赞助人 引出离散二元分布定义

**定义3.4.2 离散联合分布** 设和是随机变量,并考虑有序对.如果数量有限或者可数,则我们说和有离散联合分布.

**定理3.4.1** 假设两个随机变量和各自有离散分布.则和有离散联合分布.

**定义3.4.3** 联合概率函数p.f. 和的联合概率函数或联合p.f.被定义为函数,使得对于平面中的每个点,

**定理3.4.2** 设和有离散联合分布.如果不是的可能值,则.同样

最后,对于有序对的每一个集合,

**例题3.4.3** 通过概率表指定离散联合分布

**例题3.4.4** 水电需求,引出连续联合分布

**定义3.4.4 连续联合分布/联合p.d.f./支持** 设和是两个(连续)随机变量,如果存在一个定义在整个平面的非负函数,该平面的任意子集满足

如果积分存在,则和有连续联合分布.函数被称为和的联合概率密度函数(简称联合p.d.f).闭集被称为的支持(或分布).

**例题3.4.5** 水电需求 给出水电随机变量的严格分布.

**定理3.4.3** 一个联合p.d.f.必须满足下列两个条件:

和

**定理3.4.4** 对于每一个在平面上的连续联合分布,下面两个定理成立:

1. 平面上的每一个点,每一组无穷点序列.其概率可0.
2. 设是定义在上的一维连续实函数.则集合和的概率为0.

**证明**: 根据定义3.4.4,可以通过对联合p.d.f进行积分来找到连续联合分布分配给平面指定区域的概率.如果该区域是单个点,则积分将为0.通过概率公理3,点的任何可数集合的概率也必须为0.平面上连续函数图上两个变量的函数的积分也为0.

**例题3.4.5** 不存在连续联合分布 从定理3.4.4的(2)得出,位于平面中每个指定直线上的概率为0.如果具有连续分布并且,则和具有连续分布,但位于直线上的概率为1.因此,和不能具有连续的联合分布.(**没有看懂该例题**)

**例题3.4.7** 计算标准化系数

**例题3.4.8** 对**例题3.4.7**的扩展

**例题3.4.9** 计用几何方法确定联合p.d.f.

**例题3.4.10** 临床试验 **例题2.1.14**的扩展

**定义3.4.5 联合p.f./p.d.f.** 设和是随机变量,其中是离散的并且是连续的.假设存在一个定义在平面上的函数,对于实数子集中每一对和,

如果该积分存在.则函数被称为和的联合p.f./p.d.f..

所有的联合p.f./p.d.f必须满足两个条件.即公式必须满足

**注意:一般集合的概率** 对于实数对的一般集合,我们可以利用和的联合p.f./p.d.f.来计算.对每一个,设.则

如果所有的积分存在.此外,对每一个,设.则

如果该积分存在.

**例题3.4.11** 联合p.f./p.d.f.的应用

**例题3.4.12** 临床试验 **例题3.4.10**的解答(**比较重要,联合p.d.f.的给出比较难理解**)

**例题3.4.13** 一个复杂的联合分布 (**比较难懂**)

**定义 3.4.6 联合(累加)分布函数c.d.f.** 随机变量和的联合分布函数或联合累加分布函数(联合c.d.f)定义为函数,使得和的所有值(),

如果任意随机变量和的联合c.d.f.是函数,那么位于平面上的变量对的概率则位于特殊的矩形区域内,对于给定的数值和,

**定理3.4.5** 如果和具有联合c.d.f..则单随机变量的c.d.f.可以从导出.类似的,的c.d.f.等于,其中.

最后,如果和具有连续联合分布,则在的联合c.d.f.是

而联合p.d.f.可以从联合c.d.f.通过下列关系导出

假设二阶导数存在.

**例题3.4.14** 一从联合c.d.f计算联合p.d.f. 对定理3.4.5的详细说明,需要仔细阅读.

**例题3.4.15** 水电需求 对定理3.4.5的详细说明

**3.5 边际分布** 2020年2月13日12点20分

**例题3.5.1** 水电需求 对**例题3.4.15**的进一步阐述

**定义3.5.1 边际c.d.f./p.f./p.d.f.** 假设和具有联合分布.根据定理3.4.5导出的的c.d.f.被称为的边际c.d.f..类似的,与边际c.d.f.相关联的p.f.或p.d.f.被称为的边际p.f.或边际p.d.f..

**定理3.5.1** 如果和具有离散联合分布,其联合p.f.为,则的边际p.f.是

相似地,的边际p.f.是.(**证明过程需要理解**)

**例题3.5.2** 从概率表计算边际p.f. **例题3.4.3**的扩展

**定理3.5.2** 如果和具有连续联合分布,其联合p.d.f.为,则的边际p.d,f.是

相似地,的边际p.d.f.是

(**证明过程需要理解**)

**例题3.5.3** 根据连续联合p.d.f.推导边际p.d.f. **例题3.4.8**的扩展

**定理3.5.3** 设是和的联合p.f./p.d.f.,其中是离散的,是连续的.则的边际p.f.是

的边际p.d.f.是

**例题3.5.4** 从联合p.f./p.d.f.推导边际p.f.或边际p.d.f.

**例题3.5.5** 边际和联合分布

**例题3.5.6** 水电需求 对**例题3.4.15**和**例题3.5.1**的进一步阐述引出独立随即变量的关系

**定义3.5.2 独立随机变量** 如果两个实数集和使得和是事件,则随机变量和是独立的,

换句话说,设为发生或不发生仅取决于值的任意事件(例如),设为发生或不发生仅取决于值的任意事件(例如).当且仅当和是所有此类事件的独立事件时,和才是独立随机变量.

如果和是独立的,那么对于所有实数和,必须为

**定理3.5.4** 设和的联合c.d.f.为,设的边际c.d.f.为,设的边际c.d.f.为.则和是独立的当且仅当对所有的和满足.

**定理3.5.5** 假设和是随机变量且具有联合p.f.,p.d.f.,或p.f./p.d.f..则和是独立的当且仅当对于和可以表达成下列形式:

其中是的非负函数,是的非负函数.(**证明过程需要理解**)

**推论3.5.1** 两个随机变量和是独立的当且仅当下列因式分解对所有实数和成立:

**独立随机变量的意义**: 我们在定义3.5.2中给出了独立随机变量的数学定义,但尚未对独立随机变量的概念给出任何解释.由于独立事件和独立随机变量之间的紧密联系,独立随机变量的解释应与独立事件的解释密切相关.如果了解到其中一个事件不会改变另一个事件的发生概率,我们会将两个事件建模为独立事件.将这种想法扩展到离散随机变量是最容易的.假设和具有离散联合分布.如果对于每个得知不会改变事件的任何概率,则我们要说和是独立的.从推论3.5.1和边际p.f.的定义中,我们看到,确实当且且仅当每个使得,也就是说,研究值不会改变与相关的任何概率.当我们在3.6节中正式定义条件分布时,我们将看到对独立离散随机变量的这种解释扩展到所有双变量分布.总而言之,如果我们试图确定是否将两个随机变量和建模为独立变量,则应考虑在获知值之后是否更改的分布,反之亦然.

**例题3.5.7** 掷色子和抛硬币

**例题3.5.8** 确定随机变量在临床试验中是否独立

**例题3.5.9** 计算涉及独立随机变量的概率.

**例题3.5.10** 非独立随机变量

**定理3.5.6** 如果和具有连续联合分布.假设是一个矩形区域(可能无界)且边平行于坐标轴.则和是独立的当且仅当公式(3.5.7)对所有成立.

**例题3.5.11** 验证联合p.d.f.的因子

**注意:独立随机变量的独立函数是独立的**.如果和是独立的,则无论函数和是什么,和都是独立的.这是正确的,因为对于每一个,事件总是可以写成,其中.同样,可以写成,因此公式对于和的取自公式 对于和.

**3.6 条件分布** 2020年2月18日10点40分

**例题3.6.1** 汽车保险 引出条件分布的概念

**定义3.6.1 条件分布/p.f.** 设和具有离散联合分布,其联合p.f.为.设是的边际p.f..对每一个值使得,定义

则称为给定的的条件分布p.f..假设,则是x的条件分布.

类似地,如果是的给定值,使得,并且是给定的的条件p.f..则

**例题3.6.2** 从联合p.f.中计算条件p.f. **例题3.5.2**的扩展

**例题3.6.3** 汽车保险 **例题3.6.1**的条件概率分析

**例题3.6.4** 处理时间 引出连续分布的条件分布定义

**定义3.6.2 条件p.d.f.** 设和具有连续联合分布,其联合p.d.f.为,相应的边际分布分别为和.设每一个值使得.则在给定条件下的条件p.d.f.定义如下:

对于使得的值,我们可以随意定义,只要是p.d.f.作为的函数.

**定理3.6.1** 对每一个,定义3.6.2所示的作为的函数是一个p.d.f..

**例题3.6.5** 处理时间 **例题3.6.4**的解

**例题3.6.6** 从联合p.d.f.中计算条件p.d.f. **例题3.4.8**的扩展

**定义3.6.3 混合分布的条件p.f.或p.d.f.** 设是离散的并且设是连续的,其联合p.f./p.d.f.为.则在给定条件下的条件p.f.由公式(3.6.2)定义;在给定条件下的条件p.d.f.由公式(3.6.3)定义.

**例题3.6.7** 损坏零件 **例题3.1.9**的扩展

**定理3.6.2 分布的乘法法则** 设和是随机变量使得具有p.f.或p.d.f. 以及具有p.f.或p.d.f. .此外,假设给定时的条件p.f.或p.d.f.为;给定时的条件p.f.或p.d.f.为.则对于每个使得和每个,

其中是联合p.f.,p.d.f或p.f./p.d.f.类似地,对于每个使得和每个,

**例题3.6.7** 排队等待时间 对定理3.6.2的应用

**例题3.6.9** 损坏零件 **例题3.1.7**的扩展(**经典例题,需要细品**)

**定理3.6.3 随机变量的总概率定律** 如果是随机变量的边际p.f.或p.d.f.且给定时的条件p.f.或p.d.f.为.则的边际p.f.或p.d.f.为

如果是离散的.如果是连续的,的边际p.f.或p.d.f.为

**定理3.6.4 随机变量的贝叶斯定理** 如果是随机变量的边际p.f.或p.d.f.且是给定时的条件p.f.或p.d.f.,则给定时的条件p.f.或p.d.f为

其中从等式或获得.同样,给定时的条件p.f.或p.d.f为

其中从等式或获得且需要交换和以及下标1和2.

**例题3.6.10** 从均匀分布中选取样本 **对定理3.6.4的阐述**

**定理3.6.5 独立随机变量** 假设和是随机变量且具有联合p.f.,p.d.f.,或p.f./p.d.f. .则和是独立的当且仅当对于每一个值使得且对于每一个x值,满足

3.7 多变量分布 2020年2月18日15点00分

**例题3.7.1** 临床试验 联合分布函数/c.d.f. 个随机变量的联合c.d.f.是维空间在每个点处的函数,存在以下关系

每个多变量c.d.f.满足与先前单变量和双变量c.d.f.给出的属性相似.

**例题3.7.2** 失败时间

定义3.7.2 联合离散分布/p.f. 如果随机向量在具有有限数值或不同可能值的无穷序列,则称这个随机变量具有离散联合分布.的联合p.f.定义为函数,对于每一个点满足

用向量符号表示为

其中.

定理3.7.1 如果具有联合离散分布且联合p.f.为f,则对每一个子集,

**例题3.7.3** 临床试验,**有意义**

定义3.7.3 连续分布/p.d.f. 如果存在一个定义在上的非负函数使得每一个子集满足

且该积分存在.则称这个随机变量具有连续联合分布.函数被称为的联合p.d.f..

用向量符号表示为

其中.

定理3.7.2 如果的分布是连续的,则联合p.d.f. 可以从联合c.d.f. 根据下列关系导出

假设导数在所有点处存在.

**例题3.7.4** 失败时间 **例题3.7.2**的扩展

**例题3.7.5** 服务等待时间

**例题3.7.6** 抵达队列的时间 提出问题

定义3.7.4 联合p.f./p.d.f. 设是随机变量,其中一些具有连续联合分布,而另一些具有离散分布;它们的联合分布由函数表示,我们称之为联合p.f./p.d.f..该函数具有以下性质:通过将离散随机变量相对应的坐标值上的相加,并对与连续变量相对应的坐标进行积分,可以计算出位于子集中的概率,其中.

**例题3.7.7** 抵达队列的时间 **例题3.7.6**的解

**例题3.7.8** 临床试验 **例题3.7.1**的扩展 是对定义3.7.4,定义3.7.3和定理3.7.1的说明

导出边际p.d.f. 如果个随机变量的联合分布已知,则每一个随机变量的边际分布可以从该联合分布导出.例如,如果的联合p.d.f.为,则的边际p.d.f. 在每一个点处由下列公式给出

更一般的,个随机变量中任意个的联合分布可以从剩余的个积分中得到.

**例题3.7.9** 服务等待时间 **例题3.7.5**的扩展

导出边际c.d.f. 如果个随机变量的联合c.d.f.为.则的边际c.d.f. 可以从下列关系中获得:

更一般的,个随机变量中任意个的联合c.d.f可以从剩余的个极限中获得.

**例题3.7.10** 失败时间 **例题3.7.2**的扩展

**例题3.7.11** 失败时间 **例题3.7.2**的扩展

定义3.7.5 独立随机变量 对于任意个实数集合,如果满足

则称这个随机变量是独立的.这是定理3.5.4的一般化.

定理3.7.3 设是的联合c.d.f.,设是的单变量边际c.d.f.,其中.变量是独立的当且仅当,对于所有点满足:

定理3.7.4 如果具有连续的,离散的,或混合联合分布,其联合p.d.f.,联合p.f.,或联合p.f./p.d.f.是f,并且如果是的单变量边际p.d.f.或p.f.,则是独立的当且仅当,对于所有点满足:

**例题3.7.12** 服务等待时间 **例题3.7.9**的扩展

定义3.7.6 随机样本/i.i.d./样本大小 考虑实线上一个给定的概率分布,用p.f.或p.d.f. 表示.如果个随机变量是独立的并且它们的边际p.f.或p.d.f.为,那么这些随机变量构成该分布的一个随机样本.这样的随机变量也被称为独立同分布[independent and identically distributed],简称i.i.d.我们把随机变量的数目称为样本大小.

定义3.7.6称构成分布的一个样本空间，如果它们的联合p.f.或p.d.f. 对于所有点满足如下等式:

显然，i.i.d.样本不能有混合分布。

**例题3.7.13** 灯泡寿命 思路清奇,值得一看

**定义3.7.7** 条件p.f.,p.d.f,或p.f./p.d.f. 假设随机变量被分割成两个子向量和,其中是一个维随机向量,包含中个随机变量,是一个维随机向量包含另外个随机变量.假设的维联合p.f.,p.d.f.,或p.f./p.d.f.是并且的边际维p.f.,p.d.f.,或p.f./p.d.f为.对于给定的使得,则给定时的条件维p.f.,p.d.f.,或p.f./p.d.f. 定义如下:

公式可以被写为

**例题3.7.14** 服务等待时间 **例题3.7.9**的扩展,需要理解

**例题3.7.15** 确定边际二元p.d.f. 有点绕,需要结合定义3.7.7理解

**例题3.7.16** 服务等待时间 **例题3.7.14**的扩展

定理3.7.5 多变量总概率定律和贝叶斯理论 假设条件和符号如定义3.7.7.如果拥有连续联合分布,则的边际p.d.f.为(**该定理需要结合例题3.7.15和例题3.7.16来理解**)

并且在给定条件下的条件p.d.f为

如果具有离散的联合分布,则中的多重积分必须替换为多重求和.如果具有混合联合分布,则必须用具有连续分布的那些坐标上的积分和具有离散分布的那些坐标上的求和来代替多个积分.

**定义3.7.8 条件独立随机变量** 设是一个随机向量并且它的联合p.f.,p.d.f.,或p.f./p.d.f.为.多个随机变量在给定的情况下是条件独立的,仅当对于所有值使得,满足(**该定理需要结合例题3.7.15来理解**)

其中代表在给定条件下的条件多变量p.f.,p.d.f.,或p.f./p.d.f. 代表在给定条件下条件单变量p.f或,p.d.f.

**例题3.7.17** 临床试验 **例题3.7.8**的扩展 是对定义3.7.8的说明

**定义3.7.9 直方图.** 设是一组位于之间的数字，. 选择某些整数，将区间分割成长度为的k个等长子区间。对于每一个子区间，统计有多少数字在该子区间。设为子区间的个数, . 选择一个数字(一般或或). 绘制一个二维图像，其横轴范围从到. 对每一个子区间绘制一个长为，高为，横过第个区间的矩形条。这样的图像被称为直方图。